# Bayesian update rule leads to information theoretically optimal evolution

### Why to help those weaker than you

Tomas Ukkonen
tomas.ukkonen@iki.fi

May 27, 2016

## 1   Introduction

Evolution can be underatood as an optimization process that tries to collect information [1] as well as possible from an environment in order to increase fitness. One can think a current population as the distribution of genes $p(\mathbf{x})$ and then define/have a fitness function where outcomes are uncertain and defined as the probability of fitness $p(\mathbf{f}|\mathbf{x})$. It is then possible to use the bayes rule [2] to calculate the posterior distribution of good genes $p(\mathbf{x}|\mathbf{f})$ given the fitness outcomes $\mathbf{f}$.

$$p(\mathbf{x}|\mathbf{f}) \propto p(\mathbf{f}|\mathbf{x})p(\mathbf{x}) \tag{1}$$

$$H(\mathbf{X}|\mathbf{F}) = H(\mathbf{X}) - I(\mathbf{X};\mathbf{F}) \tag{2}$$

The bayesian update rule is information theoretically optimal meaning information is gained as fast as possible from the environment.

## 2   Information Theory

It is easy to prove equation 2 is correct from bayesian update equation 1.

$$p(\boldsymbol{x}|\boldsymbol{f})p(\boldsymbol{f}) = p(\boldsymbol{f}|\boldsymbol{x})p(\boldsymbol{x})$$
$$H(\boldsymbol{X}|\boldsymbol{F}) + H(\boldsymbol{F}) = H(\boldsymbol{X}, \boldsymbol{F})$$
$$H(\boldsymbol{X}|\boldsymbol{F}) = H(\boldsymbol{X}, \boldsymbol{F}) - H(\boldsymbol{F})$$
$$H(\boldsymbol{X}|\boldsymbol{F}) = H(\boldsymbol{X}) - I(\boldsymbol{X};\boldsymbol{F})$$

However, the reverse is not necessarily true. There may be non-bayesian methods to process data that are information theoretically optimal as well.

# 3   Conclusions

The keypoint is that in order to get good results and avoid getting stuck into local maximas, one needs to support individuals sampling low probability regions/tails of the distribution. If the individuals in low probability regions die, then the search and transfer of information doesn't happen as fast as possible and the group is likely to lose against other groups which are able to collect and process information faster. This means weaker individuals (risk takers) should be supported somehow in order to maintain the whole distribution and the inference/optimization rule. For example, stronger individuals (having enough intelligence) may gain extra information by observing and learning from mistakes of others.

However, this theoretical framework partially ignores generation of descendants through sex and is best suited for species which reproduces through mutation. Mutations happen in a distributed fashion without central control and the process is similar to MCMC sampling. Further work could involve improved modelling, computer simulations and documenting how adaptation speed and quality varies as a function of death probability for low fitness individuals, and how the use of only life expectancy as the fitness function, and not the competition, changes information transfer rates from environment.

# References

[1] MacKay D., *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.

[2] Gelman A., Carlin J., Stern H. and Rubin D., *Bayesian Data Analysis. 2nd edition.*, CRC Press, 2003.

[3] Beirlant J., Dudewicz E., Gyrfi L., van der Meulen E. *Nonparametric entropy estimation: an overview*, 2001